Distributed Systems

COMP90015 2017 Semester 1

Week 11

Tutorial 10

Things to cover today

- Quiz 6: Operating System Support
- Quiz 7: Security
- Project Questions
- Distributed File System (DFS) questions

Distributed File System (DFS) questions

1. What are some key design issues for distributed file systems?

2. Name and explain three transparencies that should
be addressed by distributed file systems.

3. What are the advantages and disadvantages of using absolute names as a
naming strategy?

4. What are the advantages and disadvantages of a naming strategy based on
mount points?

5. What are the advantages and disadvantages of using a global name space
naming strategy?

Continue

6. What are the advantages and disadvantages of the following simple distributed file system?

7. What are the advantages and disadvantages of the following distributed file system?

8. Discuss the advantages and disadvantages of storing client cache files on disk and on memory?

9. Discuss the advantages and disadvantages of a write-through client cache update policy.

10. Explain the two conditions used in NFS to determine whether a cache block at the client is valid or invalid. Explain which condition is tested first and why.

1. What are some key design issues for distributed file systems?

- Effective use of client caching to achieve performance
- Maintain consistency between multiple copies of cached files
- Recovery after client or server failure
- High throughput for reading and writing files of all sizes
- Scalability

2. Name and explain three transparencies that should be addressed by distributed file systems.

Access transparency

• Client programs don't know if the file is local of remote

Location transparency

• Client programs don't know where the file is stored

• Files can be relocated without changing their pathname

Mobility transparency

• Neither client programs nor system administration tables in client nodes need to be changed when files are moved

2. Name and explain three transparencies that should be addressed by distributed file systems.

Performance transparency

• Maintain acceptable performance while the load on the service varies within a specified range

Scaling transparency

• Service can be expanded without loss of performance

3. What are the advantages and disadvantages of using absolute names as a naming strategy?

- **Absolute name**
- provides a complete address to a file including both the server and path names: <machine name: path name>
- **Advantages**
- Trivial to find a file once the name is given
- No additional state must be kept since each name is self contained (No global state)
- Greater scalability
- Easy to add and delete new names
- **Disadvantages**
- No location transparency
- File is location dependent and cannot be moved
- Less resilient to failure

4. What are the advantages and disadvantages of a naming strategy based on mount points?

- **Mount Points (used by Sun's Network File System - NFS)**
- The client machine creates a set of "local names" which are used to refer to remote locations: mount points
- At boot time, the local name is bound to the remote name
- The operating system must maintain a table to maintain the mapping of what server and path are mapped to each mount point

- **Advantages:**
- Names do not contain information about the file location
- Remote location can change between reboots

- **Disadvantages:**
- Hard to maintain
  - What happens when machines fail?
  - What happens when files are migrated?
- Can lead to confusion since two different local names may map to the same file on a remote system

5. What are the advantages and disadvantages of using a global name space naming strategy?

- **Global Name Space**
- All nodes have an identical name space: the path and name of a file on one machine will be the same on every other machine, regardless of where the file is actually stored
- Dedicated file servers
- Client contacts one of the servers and receives the layout of the distributed file system
- When a user accesses a file, the server sends a copy of it to the client machine where it is cached
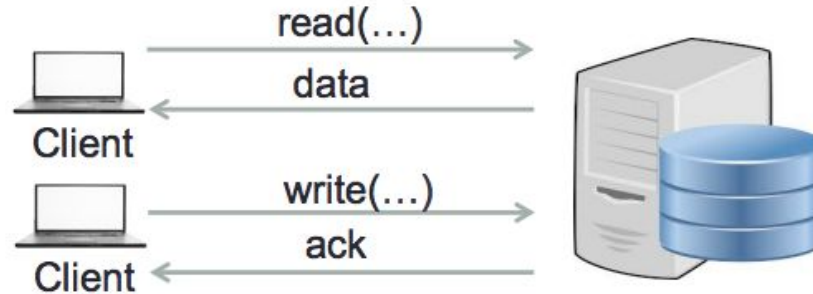
- **Advantages**
- Location transparency
- Naming is consistent across all clients
- Storage servers are able to seamlessly move files around because clients always contact the server to learn where files are located
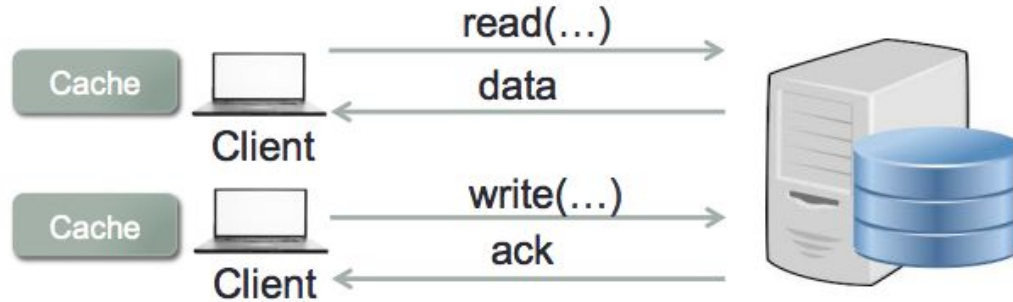
- **Disadvantages**
- Files are cached by clients = challenges in keeping files consistent
- Can lead to performance problems, particularly when the scale of the system grows

6. What are the advantages and disadvantages of the following simple distributed file system?

- RPC to access file system calls
- No client/local caching
- Advantages
- Consistent view of file system
- Disadvantages
- Performance (network access, server becomes potential bottleneck)

7. What are the advantages and disadvantages of the following distributed file system?

- Cache files on clients and perform local file system operations
- Advantages
- Local operations = better performance
- Disadvantages
- What happens when the client fails?
- Where should the client store the cached files?
- Difficult to keep local copy consistent with remote copy

8. Discuss the advantages and disadvantages of storing client cache files on disk and on memory?

- Disk
  - Advantages:
    - Safer if client fails (more reliability)
  - Disadvantages:
    - Slower than just keeping it in local memory
    - Requires client to have a disk
- Memory
  - Advantages:
    - Quick access time
    - Works with diskless workstations
- Disadvantages:
  - Does not tolerate node failure well
  - Limited cache size

9. Discuss the advantages and disadvantages of a write-through client cache update policy.

• **Write-through: All writes are propagated to the server and the server's copy of the file is updated immediately**
• **Advantages**
  • High reliability since writes are immediately written to the server
• **Disadvantages**
  • Reduced performance because all writes not only need to be made to the local cache, but transmitted over the network and acknowledged by the server
• **The cache provides no benefit for write requests (identical to doing all writes remotely), but it can still improve performance for reads**

10. Explain the two conditions used in NFS to determine whether a cache block at the client is valid or invalid. Explain which condition is tested first and why.

- **A cache block is said to be valid at time T if:**
  - The cache block is 'fresh enough': $T-Tc < t$ (Tc = time when cache entry was last validated)

Or

  - The time when the file was last modified recorded by the client and the server match: $Tmclient = TmServer$
- **The first validity condition can be evaluated without access to the server**